

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effects of parenting interventions for at-risk parents with infants: A systematic review and meta-analyses
AUTHORS	Rayce, Signe; Rasmussen, Ida; Klest, Sihui; Patras, Joshua; Pontoppidan, Maiken

VERSION 1 - REVIEW

REVIEWER	Benzies, Karen M. University of Calgary CANADA
REVIEW RETURNED	31-Jan-2017

GENERAL COMMENTS	<p>Thank you for the opportunity to review this interesting and well-written manuscript of a well-designed and executed systematic review with meta-analyses. This manuscript will make an important contribution to the field.</p> <p>Abstract: under Design, missing a word "grey literature"</p> <p>Background arguments are strong, well supported and lead to a logical conclusion to conduct the review.</p> <p>Page 6 line 4 use past tense "included"</p> <p>Please explain why CINAHL was not included in the databases searched.</p> <p>May I suggest changing from the term "caretakers" to "caregivers"?</p> <p>The forest plots presented on pages 46-48 require additional information about the outcome.</p> <p>Overall, this is an excellent manuscript.</p>
-------------------------	--

REVIEWER	Chris Rossiter Faculty of Health University of Technology Sydney Australia
REVIEW RETURNED	21-Feb-2017

GENERAL COMMENTS	<p>Review article BMJ Open</p> <p><i>The effects of parenting interventions for at-risk parents with infants: a systematic review and meta-analysis.</i></p> <p>Rayce et al</p> <p>This manuscript reports a systematic review of studies of interventions aimed to promote important outcomes for child development and parent-child relationships amongst infants in</p>
-------------------------	---

	<p>families experiencing psychosocial risk factors. This topic is of significant interest to clinicians, researchers and parents of young children. The review and meta-analysis are conducted carefully and reported thoroughly.</p> <p>Below are some relatively minor suggestions for further refinement or clarification:</p> <ul style="list-style-type: none"> • Page 7, Table 1 – population exclusions. The authors later explain why young parents were excluded as a potential indicator of risk. However, it would be useful to understand why the other factors were excluded from the review process. Why were the chosen indicators included? • Page 9, line 34 – studies with more than one outcome measure. Could the authors provide further details of the assumptions and process for pooling data from different tools into a combined measure? • Page 13, main para – description of Interventions. This is slightly confusing and may benefit from some clarification of definitions. How do ‘individual sessions alone’ differ from ‘individual sessions’ or ‘individual home visits’? Is it due to the setting of the contact? Reference 43 is included in both ‘individual home visits’ and ‘home visits + groups’, but reference 33 is not. Reference 57 is included in the text under HV, but not in the table; it should probably have been excluded according to the criteria in Table 1. • Page 16, first para – description of Outcomes. Similarly, there appear to be a few small errors in the reporting of outcome measures for various studies. For instance, reference 52 is included in Table 5 on parent/child relationship outcomes. Reference 51 is not included in either Table 4 or 5 reporting long-term outcomes. (If, however, this is actually the same study as references 49 and 50, the total number of studies reporting long-term follow up should be three.) • Page 22, line 6 – results in online Figure 4. The in-text effect is reported slightly differently from that in the figure. • Page 22, line 20 – results from individual studies. The odds ratio refers to results reported in Table 4 on socio-emotional problems not behavioural, using DECA. • Page 26, para 3 – maternal sensitivity. These follow-up results are not reported in Table 5. • Page 29, para 2 – variation in studies. While the Discussion addresses the potential impact of the varying interventions studied and the limited information on their implementation in the meta-analysis, it does not address the potential impact of combining the differing tools for measuring outcomes. In particular, it may be valuable to consider the inclusion of data from some ‘homemade’ measures of parent-child relationship. • Page 32 – Acknowledgements. Is there a reason for not including the research assistants’ names?
--	--

	<ul style="list-style-type: none"> • Page 33 – References. #3 appears to have an EndNote problem. • Page 41, Figure 3 – Meta-analysis of parent-child relationships. Table 5 does not include the effects data from Velderman under 'parent-child relationships'. The figures included in this meta-analysis are listed in Table 5 as being on maternal sensitivity. • Pp 44-48 – online figures. The absence of labels on the online figures has contributed to some confusion, as the data do not seem to correspond to the titles listed in the contents on page 38. For instance the first figure (p44) is listed as reporting internalising behaviour, but appears to include the data on externalising behaviour. The second figure (p 45) is listed as externalising behaviour, but has data on cognitive development. The figure on p46 appears to present long-term follow data. The figure on p47 appears to be internalising behaviour data. The final figure seems to be correctly labelled as maternal sensitivity. Once rectified, this should then be checked with in-text references to these figures. <p>Overall the manuscript is well-written and comprehensive, and with minor amendments will add important evidence to the literature in this area. It will also provide valuable guidance to researchers wishing to evaluate future interventions in this area.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
K. M. Benzies
University of Calgary, CANADA

1. Abstract: under Design, missing a word “grey literature”

Thank you for pointing this out, the missing word ‘literature’ is now added to the abstract.

2. Page 6 line 4 use past tense “included”

Thanks to this suggestion, we have changed the text as suggested.

3. Please explain why CINAHL was not included in the databases searched.

Thank you for this comment. We acknowledge that CINAHL is a relevant database within the field of parenting intervention for parents of infants. The primary reason for not including CINAHL in the databases searched was that we had no access to CINAHL. However, previous research has found that Medline covers the majority of titles identified in CINAHL and that the number of academic nursing journals unique to CINAHL was small (<http://info.hsls.pitt.edu/updatereport/?p=6886>). Besides Medline, several other databases were searched including PsychINFO. PsychINFO also covers the field of child development, parent-child relationship, and psychosocial interventions, areas which may

be identified in CINAHL but are not completely covered by Medline. Combining Medline and PsychINFO with the remaining databases searched we believe the search covers the field of parenting intervention among parents of infants satisfactorily.

4. May I suggest changing from the term “caretakers” to “caregivers”?

We have changed from caretakers to caregivers as suggested in the following two sentences on page 4 in the manuscript:

For infants, severe adversity typically takes the form of caregiver neglect and physical or emotional abuse.

Most of these interventions teach caregivers specific strategies and skills that foster healthy child development with an emphasis on promoting warm and responsive caregiving.

5. The forest plots presented on pages 46-48 require additional information about the outcome.

By mistake the forest plots were uploaded in the wrong order so that they do not correspond to the captions provided for the online forest plots (on page 38). Furthermore, the captions were not shown along with the online forest plot making it difficult to identify the outcome presented in the respective forest plots. We apologize for the inconvenience this have caused when reading the manuscript. We have now uploaded the forest plots in the right order and hope that this solves the problem. If additional information is still needed we will be glad to provide this.

Reviewer: 2

Chris Rossiter

Faculty of Health, University of Technology Sydney, Australia

1. Page 7, Table 1 – population exclusions. The authors later explain why young parents were excluded as a potential indicator of risk. However, it would be useful to understand why the other factors were excluded from the review process. Why were the chosen indicators included?

Thank you for this question. Basically, we wanted to look at parenting interventions offering general parenting advice to a broader group of parents with at-risk factors. We believe that parenting interventions offered to e.g., mothers with severe mental health problems such as substance abuse, families with infants born preterm or at very low birth weight, or families with other very specific challenges may differ in important ways from general parenting programs as these interventions should target the specific risks of these groups. We have extended the text as follows on page 7-8:

We excluded studies that examined parenting interventions aimed at specific risk-groups such as teen mothers; parents with severe mental health problems; or parents with children born pre-term, at low birth weight, or with congenital diseases. Families experiencing difficulties such as these have specific needs, and interventions aimed at these groups may be more targeted when compared to parenting interventions aimed at broader, at-risk groups of parents. Since our focus was parenting interventions aimed at at-risk parents in general, we excluded studies developed for specific risk-groups.

2. Page 9, line 34 – studies with more than one outcome measure. Could the authors provide further details of the assumptions and process for pooling data from different tools into a combined measure?

Thank you for this comment. We combined separate measures for three meta-analyses: PI overall behavior, PI cognitive development and PI relationship.

The process for combining data were as follows: For each meta-analysis we looked at the measures used in the individual studies and decided if they measured the same concept.

If a study provided results from only one measure we combined the results from the individual studies in the meta-analysis if we found that they measured the same overall concept.

In some cases a study provided data on two or more measures of the same concept (e.g., DECA and SDQ that both measure aspects of socio-emotional behavior). As we did not find any of the measures superior to the others, in these cases we combined all relevant measures to one pooled effect size by taking the mean of the effect sizes (SMD). This pooled effect was used in the meta-analysis. We chose this procedure because we found that a combination of the measures may reflect a broader understanding of the concept and therefore is more appropriate than just using one of the measures.

In one case (online figure 4 meta-analysis LF behavior) we only included the SDQ score in the meta-analysis even though one study provided data from another measure. We did this because the meta-analysis included only three studies.

We have added the following text to the discussion on page 29:

The outcomes applied in the individual studies vary and most meta-analyses are based on heterogeneous measures. Although the measures vary, they do measure the same underlying construct and can therefore be meaningfully combined in the meta-analyses.

3. Page 13, main para – description of Interventions. This is slightly confusing and may benefit from some clarification of definitions. How do ‘individual sessions alone’ differ from ‘individual sessions’ or ‘individual home visits’? Is it due to the setting of the contact?

We agree that our description is confusing. We now only use the following categories: home visits, web-coaching, individual sessions and group sessions. Home visits take place in the home, individual sessions take place outside the home. We have revised table 3 and the text on page 14 as follows:

Eight studies offered individual home visits,[41–43,46–48,52–56] three studies offered individual sessions (outside the home),[44,45,49–51] one study offered group sessions,[39] one study offered web-coaching,[40] two studies combined individual sessions and group sessions,[33] and one study combined home visits and group sessions.[43]

4. Reference 43 is included in both ‘individual home visits’ and ‘home visits + groups’, but reference 33 is not.

Thank you for pointing this out. We have rephrased the text – see comment 3.

5. Reference 57 is included in the text under HV, but not in the table; it should probably have been excluded according to the criteria in Table 1.

Reference 57 is, as pointed out, not a correct reference. We have corrected this, and the reference is now replaced with the correct reference, reference 56.

6. Page 16, first para – description of Outcomes. Similarly, there appear to be a few small errors in the reporting of outcome measures for various studies. For instance, reference 52 is included in Table 5 on parent/child relationship outcomes.

Thank you for noticing this. We apologize for the small errors in the reporting of outcome measures for various studies. We have now corrected the text and tables on page 17 as follows:

Five studies reported only child development outcomes,[33,43,45,54,55] five reported only parent-child relationship outcomes,[40,41,46–48,53] and six reported both.[39,42,44,49–52,56]

7. Reference 51 is not included in either Table 4 or 5 reporting long-term outcomes. (If, however, this is actually the same study as references 49 and 50, the total number of studies reporting long-term follow up should be three.)

Yes, this is a correct. Reference 51 is the same study as reference 49 and 50. We have therefore corrected the number of studies reporting long-term follow up to three studies instead of four studies, as follows:

All studies reported a post-intervention outcome. Two studies reported an outcome at short-term follow-up,[42,46,47] two at medium-term follow-up,[33,47] and three at long-term follow-up.[33,49–51,54,55]

8. Page 22, line 6 – results in online Figure 4. The in-text effect is reported slightly differently from that in the figure.

Thank you for this comment. This is true; although it turns out that there has been a mix-up in the order of the tables when we uploaded the manuscripts and corresponding tables. We have not changed anything in the manuscript as the reported effect size is right, but we will make sure to upload the figures in the right order.

We have, however, discovered a typo in online figure 4. The effect size of Fergusson 2013 was incorrectly presented as 1.17 and not 0.17 as it should be. This has now been corrected.

9. Page 22, line 20 – results from individual studies. The odds ratio refers to results reported in Table 4 on socio-emotional problems not behavioural, using DECA.

Thanks to this suggestion, we have changed the text as follows on page 23:

... and one study found a significant positive effect on child socio-emotional development (DECA)

We noticed, that we have reported the PI observer-rated behavior outcomes from Sierau and Barlow in table 4 separately from the other outcomes presented in these two studies. To avoid confusion we have moved the observer rated behavior outcome (BRS) so that they are now presented right under the other outcomes of Barlow and Sierau.

10. Page 26, para 3 – maternal sensitivity. These follow-up results are not reported in Table 5. Thank you for the comment. We have looked carefully at the table and have highlighted where the effects sizes of maternal sensitivity can be found (in the lower part of table 5).

11. Page 29, para 2 – variation in studies. While the Discussion addresses the potential impact of the varying interventions studied and the limited information on their implementation in the meta-analysis, it does not address the potential impact of combining the differing tools for measuring outcomes. In particular, it may be valuable to consider the inclusion of data from some 'homemade' measures of parent-child relationship.

We agree that the quality of homemade measures may not be as good as validated measures. We included homemade measures in two meta-analyses: parent-child relationship (Ammanti et al and van den Boom et al.) and maternal sensitivity (Ammanti et al). Ammanti et al. use a homemade system based on attachment theory (Ainsworth) and emotional availability (Biringen). van den Boom et al. use a system based on attachment theory (Ainsworth). In both cases coders received extensive training. All the measures of parent-child relationship vary: some studies only include an overall

measure (e.g., Høivik et al. and Taylor) whereas others report only some of the subscales of a measure (e.g. Barlow et al.). As all measures (both homemade and established measures) are based on attachment theory we believe that it makes sense to combine them in a meta-analysis even though they differ. We did run sensitivity checks on the analyses and as the results did not change significantly by removing the studies with the homemade measures we decided to keep them in the meta-analyses.

We have added the following text to the discussion on page 29:

The meta-analyses of parent-child relationship and maternal sensitivity included home-made measures which could potentially affect the results, however, sensitivity analyses showed that removing these outcomes from the analyses did not substantially alter the results and we therefore kept the outcomes in the analyses.

12. Page 32 – Acknowledgements. Is there a reason for not including the research assistants' names?

Thank you for reminding us about this. We have added the names of the research assistants in the acknowledgement as follows on page 33:

...Therese Lucia Friis, Line Møller Pedersen and Louise Scheel Hjorth Thomsen for conducting the screening..

13. Page 33 – References. #3 appears to have an EndNote problem.

Thank you for pointing this out. We have now fixed the problem

14. Page 41, Figure 3 – Meta-analysis of parent-child relationships. Table 5 does not include the effects data from Velderman under 'parent-child relationships'. The figures included in this meta-analysis are listed in Table 5 as being on maternal sensitivity.

Thank you for directing our attention to this. Velderman provides information on maternal sensitivity. This outcome should also be included under 'parent-child relationship' as it is in the meta-analysis on PI parent-child relationship. We have now added the result of Velderman under parent-child relationship in table 5. We have furthermore removed the separate section on maternal sensitivity in table 5 since all these effect sizes are already provided under 'parent-child relationship'. We have done this in order to be consistent in table 4 where internalizing and externalizing (being a part of child behavior) are not reported in separate sections even though separate meta-analyses were conducted for these sub-outcomes.

15. Pp 44-48 – online figures. The absence of labels on the online figures has contributed to some confusion, as the data do not seem to correspond to the titles listed in the contents on page 38. For instance the first figure (p44) is listed as reporting internalising behaviour, but appears to include the data on externalising behaviour. The second figure (p 45) is listed as externalising behaviour, but has data on cognitive development. The figure on p46 appears to present long-term follow data. The figure on p47 appears to be internalising behaviour data. The final figure seems to be correctly labelled as maternal sensitivity. Once rectified, this should then be checked with in-text references to these figures.

Thank you for pointing this out. As mentioned previously, we discovered that the forest plots were uploaded in the wrong order and without labels. We apologize for the confusion this has caused and have now uploaded the forest plots in the correct order.

VERSION 2 – REVIEW

REVIEWER	Chris Rossiter Faculty of Health University of Technology Sydney Australia
REVIEW RETURNED	13-Mar-2017

GENERAL COMMENTS	The authors have addressed previous comments very thoroughly and thoughtfully, which has helped clarify a number of specific queries.
-------------------------	---

	The manuscript is a most valuable addition to the literature in this area and an excellent example of a well-conducted and well-written systematic review and meta-analysis.
--	--

REVIEWER	Adrian Barnett Queensland University of Technology Australia
REVIEW RETURNED	15-May-2017

GENERAL COMMENTS	<p>This was an easy paper to read and review.</p> <p>The discussion currently leans to a dichotomous argument of what is "statistically significant" or not (e.g., page 28, line 52), but statistical significance is just one part of the evidence. Some of the confidence intervals from the meta-analysis only just include zero and a Bayesian perspective would give a much higher probability to the alternative than the null. A discussion that included the public health significance would be useful. Presumably even small effect sizes could be worthwhile here given the number of mothers and children at risk.</p> <p>The statement that "teen mothers are not yet fully developed" (page 30) is pejorative and is a sweeping statement that does not account for those teen mothers who do an excellent job.</p> <p>Minor comments</p> <ul style="list-style-type: none"> - Page 8, Add citation or web link for Eppi-Reviewer 4. - Page 9, Add citation or web link for Meta-Analysis Effect Size Calculator. - Page 10, last paragraph, add the information on who short-term, mid-term, etc, times were classified. This is currently given later in the results (page 17). - Page 11, typo "Cis" - Table 3, define TAU in table footnote - Page 21, line 48, use "statistically" in front of "significant" - Page 26, whole numbers should be enough accuracy for reporting the I-squared statistics - Page 27, line 54, give I-squared value - Page 29, "home-made measures" is an odd phrase, does it mean measures that have not been validated?
-------------------------	---

REVIEWER	Annalisa Perna IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Department of Renal Medicine, Clinical Research Center for Rare Diseases "Ado e Cele Daccò", Ranica, Bergamo, Italy
REVIEW RETURNED	30-May-2017

GENERAL COMMENTS	The Authors undertook a systematic review and meta-analysis based on 3142 participants enrolled in 16 randomized or quasi-randomized trials in order to assess whether parenting interventions offered to at-risk families with infants aged 0-12 months may improve child behavior and parent-child relationship. They found a small, but significant improvement in short-term child behavior. The meta-analysis on the parent-child relationship showed a positive effect of the intervention, although substantial heterogeneity may be
-------------------------	---

	<p>present.</p> <p>Three main issues are affecting the present systematic review:</p> <p>a) absence of a prospectively registered protocol: it is virtually impossible to clearly identify a priori from post-hoc analyses.</p> <p>b) high degree of heterogeneity in the definition of 'risk' (ranging from 'insecure attachment' to 'low income', see Table 2), as far as in study participants, interventions (see Table 3) and outcomes using standardized vs non-standardized scales (see Table 5). Unfortunately the low number of included trials didn't allow subgroup analyses to better investigate heterogeneity.</p> <p>c) The assessment of reporting biases - which in this field of investigation are likely even more pronounced than in other fields - is not adequately addressed. Please provide a figure showing the relationship between effect size and standard error using funnel plots (Egger 1997). An asymmetry may suggest publication bias due to systematic differences between small and large studies. In this systematic review there was a considerable variation in sample size of the individual studies, ranging from 40 to about 800 participants. In addition a table reporting characteristics of the excluded studies together with the reason(s) for exclusion should be provided.</p> <p>Further several other aspects should be clarified. More specifically:</p> <ul style="list-style-type: none"> - What about the role of socioeconomic status or ethnicity? - In Table 2 for the Bridgeman et al study (ref 45) the Authors were not able to retrieve the exact number of participants: how was it possible to include ref 45 in the analysis without knowing this crucial information? - The Authors assessed that they included 16 individual studies (page 11 line 11): however 15 studies only are listed in Table 2. Please clarify. - It may happen that a clustering effect arises when interventions are provided in group sessions (e.g. Kaminski et al ref 34, Katz et al ref 44). How was this potential issue referred to the unit of analysis (individual vs cluster) addressed? - It is difficult for the reader to understand which measures of outcome are considered standardized and which not. - Primary results - i.e. child behavior and parent-child relationship - don't appear robust enough. The Authors assessed the analysis showed a small but significant effect in child behavior in favor of the intervention group (pag 21, line 18). It is however unclear the exact meaning of significant: statistically? clinically? On the other hand a more pronounced effect was observed in parent-child relationship, but a substantial heterogeneity was found. Removing van den Boom et al study in a sensitivity analysis not only heterogeneity but effect size also remarkably decreased. The Authors should attempt to explain the reasons why the above study likely differ from the others: maybe the use of non-standardized scales? - How did the Authors deal with relevant missing/dropout data? What about the level of attrition in the 16 included studies? What about the use of intention-to-treat principle in the primary analyses of the individual studies? - Figures 2 and 3 and online Figures 1-5 should specify that the 95% Confidence intervals refer to a random effects model. Furthermore the number of participants in each group of intervention, in each individual study should be specified.
--	---

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3

Adrian Barnett

Queensland University of Technology, Australia

Please state any competing interests or state 'None declared': None declared

1. The discussion currently leans to a dichotomous argument of what is "statistically significant" or not (e.g., page 28, line 52), but statistical significance is just one part of the evidence. Some of the confidence intervals from the meta-analysis only just include zero and a Bayesian perspective would give a much higher probability to the alternative than the null. A discussion that included the public health significance would be useful. Presumably even small effect sizes could be worthwhile here given the number of mothers and children at risk.

This is helpful feedback and we have added to the discussion a section to help highlight these issues. On page 30, we added the following to address the statistical significance of some of the subscales:

The tests for the child behavior subscales internalizing and externalizing narrowly included the zero value within in the 95% CIs (-0.03 to 0.33 and 0.00 to 0.30, respectively). These values suggest that similar studies to those in this review would likely produce small but positive effects. Because these analyses are based on three studies, there is a certain degree of uncertainty regarding the CIs reported. A larger sample of studies may be necessary to conclusively determine the significance of these results.

We also added a brief discussion of the small effect sizes on p.29. New text is emphasized in bold here:

The meta-analyses showed the most pronounced effect sizes for parent–child interaction and maternal sensitivity, whereas the effects on child behavior and cognitive development were either small or not significant, however, small effect sizes can have meaningful impact on population-level outcomes (Embry, 2011). The non-significant outcomes for internalizing and externalizing behaviors were also small, but may be clinically relevant for large, at-risk populations."

2. The statement that "teen mothers are not yet fully developed" (page 30) is pejorative and is a sweeping statement that does not account for those teen mothers who do an excellent job.

Thank you for pointing out this wording error. We did not mean to imply that teen mothers are not good parents only that they are at a different developmental phase than adults and therefore, the parenting programs and outcomes often differ significantly between teen mothers and adult mothers (teen mothers often show better outcomes than adult mothers, but the interventions also differ greatly between the two groups). The manuscript text has been updated on p 31 as follows:

Although teen mothers are an at-risk group due to their age, and they often face additional risk factors such as poverty, low education, and single parenthood, we have not included them in this review. We believe this is the appropriate method because teen mothers are a distinct group requiring targeted care that is developmentally appropriate for their stage in life. We consider the narrower focus on adult mothers to be a strength, because the interventions aimed at adult mothers most often differ considerably from interventions for teen mothers; this specificity reduces heterogeneity in study outcomes that are often present between the teen and adult interventions.

Minor comments

3. Page 8, Add citation or web link for Eppi-Reviewer 4.

We have added the reference

4. Page 9, Add citation or web link for Meta-Analysis Effect Size Calculator.

We have added the reference

5. Page 10, last paragraph, add the information on who short-term, mid-term, etc, times were classified. This is currently given later in the results (page 17).

Thanks to this suggestions, we have clarified the text on p 10 as follows:

Results were summarized for child development (behavior, cognitive development, psychomotor development, and communication/language) and parent–child relationship (relationship, sensitivity, and attachment classification) outcomes for the following assessment times: post-intervention (PI- immediately after intervention ending), short-term (ST - less than 6 months after intervention ending), medium-term (MT - 7–12 months after intervention ending), and long-term (LT - more than 12 months after intervention ending) follow-up.

6. Page 11, typo "Cis"

We have corrected to Cis as suggested

7. Table 3, define TAU in table footnote

We have added “TAU: Treatment as Usual” in the footnote as suggested

8. Page 21, line 48, use "statistically" in front of "significant"

We have added “statistically” as suggested

9. Page 26, whole numbers should be enough accuracy for reporting the I-squared statistics

We have changed accordingly

10. Page 27, line 54, give I-squared value

We have added the I2 value as suggested

11. Page 29, "home-made measures" is an odd phrase, does it mean measures that have not been validated?

We use the phrase “home-made” for measures that authors developed as part of the trials. None of these measures seem to be validated. We agree that “home-made” is awkward and have changed it to “in-house” in the text and included an explanation of what this means. Three of the studies use these measures for parent-child relationship outcomes:

Ammanti 2006 use the Scales of Mother-Infant Interactional System that one of the authors developed prior to the trial based on “attachment theory (Ainsworth et al., 1978) on recent perspectives about mother–infant interactional dyads as mutually regulated systems (Tronick, 1989; Tronick & Cohn, 1989), and on emotional availability (Biringen, 2000; Biringen & Robinson, 1991)”

Bridgeman 1981 Louisiana: use a Positive Language outcome that was coded based on a video recording obtained in the waiting room where they counted positive and negative interactions.

Van den Boom 1994 use child and parent interactive behavior that were created based on factor analyses of a larger set of items.

We have clarified the text on p 30 as follows:

The meta-analyses of parent-child relationship and maternal sensitivity included in-house measures, that is, measures developed by the evaluators that have, to our knowledge, not been formally validated. This could potentially affect the results, however, sensitivity analyses showed that removing these outcomes from the analyses did not substantially alter the results, therefore, we kept the outcomes in the analyses.

Reviewer: 4

Annalisa Perna

IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Department of Renal Medicine, Clinical Research Center for Rare Diseases "Ado e Cele Daccò", Ranica, Bergamo, Italy Please state any competing interests or state 'None declared': None declared

Three main issues are affecting the present systematic review:

1. absence of a prospectively registered protocol: it is virtually impossible to clearly identify a priori from post-hoc analyses.

We agree that it would be ideal with an a priori protocol. We planned the project to examine a wide array of parenting interventions for parents with infants. We did a broad search and then decided on which specific target groups we would do reviews of. Therefore we did not register a protocol up front. We already published the review of universal parenting interventions (Pontoppidan et al. 2016) and are finishing up a review of parenting interventions for mothers with depression. All three reviews follow the same structure.

2. high degree of heterogeneity in the definition of 'risk' (ranging from 'insecure attachment' to 'low income', see Table 2), as far as in study participants, interventions (see Table 3) and outcomes using standardized vs non-standardized scales (see Table 5). Unfortunately the low number of included trials didn't allow subgroup analyses to better investigate heterogeneity.

Although there is heterogeneity in the risk factors, all families exhibit risk factors (e.g., insecure attachment, poverty, low education, and living in deprived areas) that are highly correlated with each other and are well known predictors of detrimental outcomes for families. Although, these risk factors differ, the interventions for each are very similar and the outcomes of the risk factors for families are often the same (e.g., child problem behavior). In practice, families with these differing risk factors are frequently treated together. We therefore believe that it is correct to include families with different risk factors but receiving relatively similar interventions in the review.

3. The assessment of reporting biases - which in this field of investigation are likely even more pronounced than in other fields - is not adequately addressed. Please provide a figure showing the relationship between effect size and standard error using funnel plots (Egger 1997). An asymmetry

may suggest publication bias due to systematic differences between small and large studies. In this systematic review there was a considerable variation in sample size of the individual studies, ranging from 40 to about 800 participants. In addition a table reporting characteristics of the excluded studies together with the reason(s) for exclusion should be provided.

We acknowledge the risk that outcomes with significant effects are more likely reported compared non-significant outcomes and that this may lead to reporting biases. Cochrane Handbook do, however, advise against using funnel plots when the number of studies included in the meta-analysis is smaller than 10. The following is taken from Cochrane 10.4.3.1 Recommendations on Testing for Funnel Plot Asymmetry, "As a rule of thumb, tests for funnel plot asymmetry should be used only when there are at least 10 studies included in the meta-analysis, because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry." Since all of our meta-analyses include fewer than 10 studies we do respectfully, not believe that adding funnel plots would not produce meaningful information.

Further several other aspects should be clarified. More specifically:

4. What about the role of socioeconomic status or ethnicity?

We recognize that socioeconomic status and ethnicity are important factors. All families included in the review exhibit risk factors that are related to low socioeconomic status. Two studies specifically recruited families from an ethnic minority: Katz recruited African American mothers and Mendelsohn recruited Latina mothers. Both African American mothers and Latina mothers are over-represented in the group of at-risk families with low socioeconomic status in the US where these two studies originate. The challenges that at-risk families face are relatively similar independent of their ethnicity and the interventions are very similar. We have therefore not looked further into the ethnic background of the participants.

5. In Table 2 for the Bridgeman et al study (ref 45) the Authors were not able to retrieve the exact number of participants: how was it possible to include ref 45 in the analysis without knowing this crucial information?

The Bridgeman study is a report of evaluating the PCDC intervention in three different areas. The three evaluations are different and only the New Orleans study was included in our review. As the publication is relatively old (1981) it does not contain as much information as more recent studies following e.g. CONSORT guidelines. Unfortunately, we are not able to find any information on how many families were initially recruited into the study when the children were 2 months old. We recognize that this is problematic which is also reflected in the risk of bias assessment of the study. We do have information on the exact n of participants in all the outcome analyses. The Bridgeman study is included in the meta-analyses of parent-child relationship and sensitivity. Sensitivity analyses show that excluding the Bridgeman study does not alter the results. The Bridgeman study is also included in the meta-analysis of cognitive development. Removing the study from the meta-analysis makes the effect size smaller, but as it is not significant it does not change the conclusion.

6. The Authors assessed that they included 16 individual studies (page 11 line 11): however 15 studies only are listed in Table 2. Please clarify.

We are aware that this can be confusing. The Kaminski et al. 2013 study represent two trials (LA & Miami) and is handled as two studies when reporting results. As the overall characteristics of the two sub-trials are identical we did not split the study into two in table 2.

7. It may happen that a clustering effect arises when interventions are provided in group sessions

(e.g. Kaminski et al ref 34, Katz et al ref 44). How was this potential issue referred to the unit of analysis (individual vs cluster) addressed?

To our knowledge clustering was not touched upon in the studies. All participants were individually randomized to intervention or control. We agree that there may be an effect of having the same therapist treating a group. Standard methods of dealing with clustering do, however, require a relatively large number of clusters (>50) and analysis with fewer clusters could potentially result in biased inference (Mackinnon and Webb 2016). It is not standard practice within our field to account for clustering and almost no studies do so, this is likely due to sample size restrictions within the scope of the studies.

8. It is difficult for the reader to understand which measures of outcome are considered standardized and which not.

We are sorry that this is confusing. We have not been able to examine which outcomes have been properly validated and can therefore not provide exact information on this.

9. Primary results - i.e. child behavior and parent-child relationship - don't appear robust enough. The Authors assessed the analysis showed a small but significant effect in child behavior in favor of the intervention group (pag 21, line 18). It is however unclear the exact meaning of significant: statistically? clinically? On the other hand a more pronounced effect was observed in parent-child relationship, but a substantial heterogeneity was found. Removing van den Boom et al study in a sensitivity analysis not only heterogeneity but effect size also remarkably decreased. The Authors should attempt to explain the reasons why the above study likely differ from the others: maybe the use of non-standardized scales?

Thank you for pointing this out. According to this comment and to reviewer 3s first comments we have added clarification to the discussion.

10. How did the Authors deal with relevant missing/dropout data? What about the level of attrition in the 16 included studies? What about the use of intention-to-treat principle in the primary analyses of the individual studies?

We agree that these are important issues. They are all considered as an integrated part of the risk of bias assessment.

11. Figures 2 and 3 and online Figures 1-5 should specify that the 95% Confidence intervals refer to a random effects model. Furthermore the number of participants in each group of intervention, in each individual study should be specified.

We have added CI and number of participants in each group to the figures as suggested.

VERSION 3 – REVIEW

REVIEWER	Adrian Barnett Queensland University of Technology Australia
REVIEW RETURNED	06-Jul-2017
GENERAL COMMENTS	The authors have answered my queries. I'm not familiar with this content area, but the authors appear to have done a thorough review and meta-analyses that will be of interest to the field.

	<p>Minor comments</p> <ul style="list-style-type: none"> - Top of page 28, the second mean tau value is missing. - "The effect did, however, approach insignificance." As I mentioned before, I don't think it is useful to have a strict significant vs not significant threshold at a p of 0.05. It might be better to say something like, "but the confidence interval widened". - Typo in Conclusion paragraph "statically"
--	--

VERSION 3 – AUTHOR RESPONSE

Reviewer: 3

Adrian Barnett

Queensland University of Technology, Australia

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

The authors have answered my queries. I'm not familiar with this content area, but the authors appear to have done a thorough review and meta-analyses that will be of interest to the field.

Minor comments

1. Top of page 28, the second mean tau value is missing.

We have added the tau value

2. "The effect did, however, approach insignificance." As I mentioned before, I don't think it is useful to have a strict significant vs not significant threshold at a p of 0.05. It might be better to say something like, "but the confidence interval widened".

We have changed the text on p 28 as follows:

For the parent–child relationship the effect was almost unchanged when Bridgeman et al. (1981) and Høivik et al. (2015) were removed, but the confidence interval widened ($d=0.47$; 95% CI: 0.00 to 0.95).

3. Typo in Conclusion paragraph "statically"

We have corrected the typo